



Lecture questions

In this exercise you will answer questions related to this week’s lecture.

Note that you can submit the answer to each question only once. The questions are designed to be easy if you have followed the lecture. Note that the questions can vary slightly between students.

Points 10 / 10

My submissions 1 / 1

Deadline Friday, 24 March 2023, 19:00

To be submitted alone

The deadline for the assignment has passed (Wednesday, 5 April 2023, 19:00).

Stop word removal

Question 1 10 / 10

While preprocessing a text, we remove stop words because they rarely ever occur in a language and therefore contribute little to the overall meaning of the text.

- ☒ False
- ☐ True

Correct!

Submit

Points 10 / 10

My submissions 1 / 1

Deadline Friday, 24 March 2023, 19:00

To be submitted alone

The deadline for the assignment has passed (Wednesday, 5 April 2023, 19:00).

Components of preprocessing

Question 1 10 / 10

Which of the following is not a component of preprocessing text for analyses?

- ☐ Noise removal
- ☒ Composition
- ☐ Normalization
- ☐ Tokenization

Correct!

Submit

Points 10 / 10

My submissions 1 / 1

Deadline Friday, 24 March 2023, 19:00

To be submitted alone

The deadline for the assignment has passed (Wednesday, 5 April 2023, 19:00).

Learning from examples

Question 1 10 / 10

Which of the following machine learning algorithms “learns rules from examples”?

- ☐ Unsupervised learning
- ☒ Supervised learning

Correct!

Submit

Points 10 / 10

My submissions 1 / 1

Deadline Friday, 24 March 2023, 19:00

To be submitted alone

The deadline for the assignment has passed (Wednesday, 5 April 2023, 19:00).

Evaluation metric

Question 1 10 / 10

Which metric of evaluating supervised machine learning models is defined as “The number of correct classifications divided by all classifications performed”?

- ☐ Precision
- ☐ Recall
- ☒ Accuracy

Correct!

Submit

Points 10 / 10

My submissions 1 / 1

Deadline Friday, 24 March 2023, 19:00

To be submitted alone

The deadline for the assignment has passed (Wednesday, 5 April 2023, 19:00).

Training/test sets

Question 1 10 / 10

Why is the labelled data typically split into separate training and test sets?

- ☐ To reduce the dimensionality of data
- ☒ To avoid overfitting the model to the data
- ☐ To reduce the data size and hence the computation time
- ☐ All of the above
- ☐ None of the above

Correct!

Submit

Points 10 / 10

My submissions 1 / 1

Deadline Friday, 24 March 2023, 19:00

To be submitted alone

The deadline for the assignment has passed (Wednesday, 5 April 2023, 19:00).

Counting

Question 1 10 / 10

Does the following statement capture what we discussed in the lecture “In its most basic form, text analysis is just about counting words. While counting is easy, choosing what to count and why is much harder.”?

- ☐ False
- ☒ True

Correct!

Submit